NPS55-78-5

NAVAL POSTGRADUATE SCHOOL

Monterey, California

REGENERATIVE SIMULATION

WITH INTERNAL CONTROLS

by

Donald L. Iglehart

and

Peter A. W. Lewis

January 1978

Approved for public release; distribution unlimited.

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral Isham Linder Superintendent	Jack R. Borsting Provost

This report was prepared by:

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS55-78-5	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Regenerative Simulation with Internal Controls		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Donald L. Iglehart and Peter A. W. Lewis		8. CONTRACT OR GRANT NUMBER(s) N00014-73-C-0086; N00014-76-C-0578; N00014-78-WR-80035
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, Ca. 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, Va. 22217		12. REPORT DATE January 1978
		13. NUMBER OF PAGES 33
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Regenerative simulation
Contral variables
Wating time processes
Variance reduction
Stable stochastic processes
Internal control variables

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We introduce a new variance reduction technique, called internal control variables, to be used in the context of regeneration simulations. The idea is to identify a sequence of control random variables, each one defined within a regenerative cycle, whose mean can be calculated analytically. These controls should be highly correlated with the usual quantities observed in a regenerative simulation. This correlation reduces the variance of the estimate for the parameter of interest. Numerical examples are included for the waiting time process of an M/M/1 queue and for several Markov chains.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REGENERATIVE SIMULATION WITH INTERNAL CONTROLS*

by

Donald L. Iglehart
Stanford University
Stanford, Ca. 94305

Peter A. W. Lewis
Naval Postgraduate School
Monterey, Ca. 93940

ABSTRACT

We introduce a new variance reduction technique, called internal control variables, to be used in the context of regeneration simulations. The idea is to identify a sequence of control random variables, each one defined within a regenerative cycle, whose mean can be calculated analytically. These controls should be highly correlated with the usual quantities observed in a regenerative simulation. This correlation reduces the variance of the estimate for the parameter of interest. Numerical examples are included for the waiting time process of an M/M/1 queue and for several Markov chains.

*This work was supported by Office of Naval Research contracts N00014-72-C-0086 (NR-047-106), N00014-76-C-0578 (NR-042-343), N00014-78-WR-80035 (NR-024-374), and National Science Foundation Grant MCS75-23607. Reproduction in whole or in part is permitted for any purpose of the United States Government.

1-30-78

REGENERATIVE SIMULATION WITH INTERNAL CONTROLS

by

Donald L. Iglehart and Peter A. W. Lewis

1. Introduction and Summary

Simulators are frequently faced with the task of estimating a parameter associated with the limiting distribution of a stable stochastic process which is being simulated. A methodology based on regenerative processes for obtaining point estimates and confidence intervals for such parameters from a single realization of the process has recently been developed in Crane and Iglehart (1974a,b), (1975a,b) and Iglehart (1975), (1976a,b), and (1977). In this paper we shall introduce a new technique, internal control variables, which can be used in conjunction with the regenerative method for obtaining additional variance reduction for the estimates.

Suppose regenerative process $\{X_t : t \geq 0\}$, which is stable in the sense that $X_t \Rightarrow X$ as $t \to \infty$ (here $\Rightarrow$ denotes weak convergence), is being simulated by generating a single realization of the process. For convenience think of this process as being either a positive recurrent Markov chain or the waiting time process for a single server queue with traffic intensity less than one. Then simulators are frequently interested in estimating $r = E\{f(X)\}$, for a given function f. The principal goal of the regenerative method is to produce a confidence interval estimate of r. The regenerative method begins by observing the pairs

of random variables $\{(Y_k,\tau_k):1 \leq k \leq n\}$, where τ_k is the length of the kth regenerative cycle and Y_k is the area under the function $f(X_t)$ in the kth cycle. Two basic facts are crucial for the regenerative method of simulating stable processes. First, the pairs $\{(Y_k,\tau_k):1 \leq k \leq n\}$ are independent and identically distributed (i.i.d.), and second $r = E\{f(X)\} = E\{Y_1\}/E\{\tau_1\}$. This last fact suggests that a natural, strongly consistent point estimate for r is $\hat{r}(n) = \bar{Y}(n)/\bar{\tau}(n)$, where $\bar{Y}(n) = n^{-1} \sum_{k=1}^{n} Y_k$ and $\bar{\tau}(n) = n^{-1} \sum_{k=1}^{n} \tau_k$. To form a confidence interval for r we can use the central limit theorem (c.l.t.)

$$\sqrt{n}\ (\hat{r}(n) - r)/(\sigma/E\{\tau_1\}) \Rightarrow N(0,1) , \tag{1.1}$$

where $N(0,1)$ denotes a mean zero, variance one normal random variable and $\sigma^2 = E\{[Y_1 - r\tau_1]^2\}$. Of course, the constant $\sigma/E\{\tau_1\}$ will have to be estimated in most simulations.

The idea behind internal control variables is to introduce a third sequence of i.i.d. random variables $\{C_k:1 \leq k \leq n\}$ which has the property that C_k is defined in terms of the kth cycle and $E\{C_k\}$ is known (or can be calculated analytically). Then another strongly consistent estimate for r is

$$\hat{r}_{CT}(n) \equiv \frac{n^{-1} \sum_{k=1}^{n} [Y_k + \beta(C_k - E\{C_k\})]}{\bar{\tau}(n)} . \tag{1.2}$$

The subscript CT is meant to denote "controlling the top" in the ratio estimator; similar estimates will be defined for "bottom control."

A c.l.t. analogous to (1.1) also holds for $\hat{r}_{CT}$ with σ replaced by σ', say. Since we are still free to select β, we choose to do so in such a way as to minimize σ'. Having done that we find that

$$(\sigma')^2 = \sigma^2[1 - \rho^2(C_1, Y_1 - r\tau_1)] .$$

To obtain significant variance reduction a control, C_1, must be found which is highly correlated with $Y_1 - r\tau_1$, a task that is not always easy.

Section 2 of the paper develops the method of internal control variables in detail and contains specific examples associated with the single server queue and Markov chain models. These ideas are then illustrated with simulation results and numerical calculations in Section 3. Concluding remarks are made in Section 4. In particular we discuss the possibility of combining internal controls with external controls to obtain further variance reduction.

2. Internal Controls

2.1. General Ideas

Let $\underset{\sim}{X} = \{X_t : t \geq 0\}$ be the regenerative process being simulated. Recall that a process is regenerative if a renewal process $\underset{\sim}{T} = \{T_n : n \geq 0\}$ is defined on the same probability space as $\underset{\sim}{X}$ and if the portions of the $\underset{\sim}{X}$ process between consecutive regeneration points, the T_n's, are i.i.d. Typically the T_n's denote the times the $\underset{\sim}{X}$ process enters a fixed state and at

these times the process starts over from scratch independent of its past history. For a formal definition of regenerative process and general background on the regenerative method of simulation analysis consult Crane and Iglehart (1975a). Let $T_0 = 0$ for simplicity and define $\tau_k = T_k - T_{k-1}$, $k \geq 1$. The portion of $\underset{\sim}{X}$ in the interval $[T_{k-1}, T_k)$ is referred to as the kth cycle and is of length τ_k. Suppose now that $E\{\tau_1\} < \infty$ and that the distribution function of τ_1 is aperiodic (e.g. its support is not contained in a set of the form $\{0,h,2h,\ldots\}$, where $h > 0$). Then subject to mild regularity conditions $X_t \Rightarrow X$ as $t \nearrow \infty$, where $\Rightarrow$ means $P\{X_t \leq x\} \to P\{X \leq x\}$ for all continuity points (x) of the limit distribution. A similar result holds when τ_1 is periodic; see Crane and Iglehart (1975a). The random variable $\underset{\sim}{X}$ is frequently thought of as the steady-state configuration of the system being simulated. Suppose f is a measurable function from the state space of $\underset{\sim}{X}$ to the real line and that we are interested in estimating $r = E\{f(X)\}$. Define the sequence of random variables (r.v.'s) $\{Y_k : k \geq 1\}$ by

$$Y_k = \int_{T_{k-1}}^{T_k} f(X_s)ds \;, \qquad k \geq 1 \; . \tag{2.1}$$

If the time parameter of $\underset{\sim}{X}$ is discrete (as in a discrete time Markov chain), the integral in (2.1) should be replaced by a sum. The regenerative stucture of $\underset{\sim}{X}$, the process being simulated, gives us the two important properties stated in Section 1: the pairs

$\{(Y_k,\tau_k): 1 \le k \le n\}$ are i.i.d. and $r = E\{Y_1\}/E\{\tau_1\}$, provided $E\{|f(X)|\} < \infty$, which we shall always assume. The c.l.t. for the ratio estimator $\hat{r}(n) = \bar{Y}(n)/\bar{\tau}(n)$ indicated in (1.1) is proved in Crane and Iglehart (1975a) and follows from the classical c.l.t. for the i.i.d. mean zero, finite variance r.v.'s $Z_k = Y_k - r\tau_k$, $k \ge 1$. We always assume $0 < \sigma^2 = E\{Z_k^2\} < \infty$. The variance of Z_1, σ^2, is also related to the variance of $\hat{r}(n)$ through the asymptotic relation (as $n \to \infty$)

$$(2.2) \qquad \sigma^2\{\hat{r}(n)\} \sim n^{-1}(\sigma^2\{Z_1\}/E^2\{\tau_1\}) .$$

For a derivation of this result see Cramér (1964, p. 354, eq. (27.7.3)). A number of point estimates and confidence intervals have been proposed for ratios such as r (see Iglehart, 1975), but the simplest conceptually and computationally are the so-called classical ones. The point estimate is $\hat{r}(n)$ and the confidence interval is

$$(2.3) \qquad \hat{I}(n) = [\hat{r}(n) - z_{1-\gamma/2}\hat{s}/\bar{\tau},\ \hat{r}(n) + z_{1-\gamma/2}\hat{s}/\bar{\tau}] ,$$

where $z_{1-\gamma/2} = \Phi^{-1}(1 - \gamma/2)$, the inverse of the standard normal distribution function, and $\hat{s}$ is the classical point estimate of σ which is constructed as follows. Let $\hat{s}_{11}$ [resp. $\hat{s}_{22}$] be the sample variance of the Y_k's [resp. τ_k's] and $\hat{s}_{12}$ the sample covariance of the Y_k's and τ_k's. Then $\hat{s}$ is defined by

$$\hat{s} = [\hat{s}_{11} - 2(\bar{Y}/\bar{\tau})\hat{s}_{12} + (\bar{Y}/\bar{\tau})^2\hat{s}_{22}]^{1/2} .$$

Introduction of the internal control variables mentioned in Section 1 is done with the hope of reducing the variance of the point estimator of r. This in turn will either reduce the number of cycles that need to be simulated for fixed precision or reduce the length of the confidence interval, (2.3), for r. The sequence of internal control variables $\{C_k : k \geq 1\}$ is defined on the same probability space supporting the process $\underset{\sim}{X}$ in such a way that C_k only depends on $\underset{\sim}{X}$ (or underlying r.v.'s defining $\underset{\sim}{X}$) in the kth cycle. This construction of the C_k's will insure that they are i.i.d. because of the basic i.i.d. structure of the cycles of a regenerative process. The control variables C_k allowed, however, are further restricted to those for which the mean, $E\{C_1\}$, is either known or can be calculated analytically. Thus one of the prices we must be prepared to pay to obtain variance reduction for our simulation is the analytical work of computing $E\{C_1\}$. Having defined the C_k's we proceed to form new ratio estimators for r. Starting with the ratio estimator $\hat{r}(n) = \bar{Y}(n)/\bar{\tau}(n)$ we can either control the Y_k's, the τ_k's, or both if we introduce a second sequence of control variables. The estimator $\hat{r}_{CT}(n)$, proposed in (1.2), involves controlling the Y_k's only. Here the subscript CT is meant to suggest "controlling the top." For convenience let (for $k \geq 1$)

$$Y_k' = Y_k + \beta(C_k - E\{C_k\})$$

and

$$Z_k' = Y_k' - r\tau_k \ .$$

Observe that the Y_k''s are i.i.d. with $E\{Y_1'\} = E\{Y_1\}$ and consequently the Z_k''s are also i.i.d. with $E\{Z_k'\} = 0$. Hence the standard c.l.t. yields (as $n \to \infty$)

$$\sum_{k=1}^{n} Z_k'/\sigma\{Z_1'\}n^{1/2} = n^{1/2}[(\bar{Y}/\bar{\tau})-r]/(\sigma\{Z_1'\}/\bar{\tau}) \Rightarrow N(0,1)$$

which, upon replacing $\bar{\tau}$ by $E\{\tau_1\}$ in the denominator (this can be justified by a continuous mapping argument), becomes

$$\frac{n^{1/2}[\hat{r}_{CT}(n) - r]}{(\sigma\{Z_1'\}/E\{\tau_1\})} \Rightarrow N(0,1).$$

The variance of Z_1', which obviously depends on β, is easily seen to be

$$\sigma^2\{Z_1'\} = \sigma^2\{Z_1\} + 2\beta \operatorname{cov}\{Z_1, C_1\} + \beta^2\sigma^2\{C_1\} .$$

Now select β so as to minimize $\sigma^2\{Z_1'\}$. This yields

$$\beta^* = - \frac{\operatorname{cov}\{Z_1, C_1\}}{\sigma^2\{C_1\}} \tag{2.4}$$

and

$$\sigma^2\{Z_1'\} = \sigma^2\{Z_1\} [1 - \rho^2\{Z_1, C_1\}] , \tag{2.5}$$

where $\rho\{Z_1, C_1\}$ is the correlation coefficient between Z_1 and C_1. If we use the c.l.t. for the estimator $\hat{r}_{CT}(n)$ as a basis for

constructing confidence intervals for r, then we want a control variable C_1 which will minimize $\sigma^2\{Z_1'\}$. From (2.5) we see that we need a large value of $\rho^2\{Z_1,C_1\}$. Since the length of such a confidence interval is proportional to $\sigma\{Z_1'\}/n^{1/2}$, to reduce the number of cycles required by a factor of four we need $|\rho\{Z_1,C_1\}| > (0.75)^{1/2} = 0.866$. To obtain a control, C_1, which is highly correlated with either Y_1 or τ_1 is relatively easy to do. However, because Y_1 and τ_1 are themselves highly correlated, it is much more difficult to find a control C_1 which is highly correlated with Z_1. In general, we shall try to find a control C_1 which mimics Z_1 but for which we can still calculate its mean, $E\{C_1\}$. As usual we try to do as much analytically as possible, before having to resort to simulation. We illustrate some possible candidates for controls in the context of GI/G/1 queues and discrete time Markov chains.

2.2. Internal controls for GI/G/1 Queues

In the GI/G/1 queue we assume the zeroth customer arrives at time $t_0 = 0$, finds a free server, and experiences a service time v_0. The nth customer arrives at time t_n and experiences a service time v_n. Let the interarrival times $t_n - t_{n-1} = u_n$, $n \geq 1$. Assume the two sequences $\{v_n : n \geq 0\}$ and $\{u_n : n \geq 1\}$ each consists of independent, identically distributed (i.i.d.) random variables (r.v.'s) and are themselves independent. Let $E\{v_n\} = \mu^{-1}$, $E\{u_n\} = \lambda^{-1}$, and $\rho = \lambda/\mu$, where $0 < \lambda,\mu < \infty$. Thus

μ has the interpretation of the mean service rate and λ has the interpretation of the mean interarrival rate. The parameter ρ is called the traffic intensity and is the natural measure of congestion for this system. We shall assume that $\rho < 1$, a necessary and sufficient condition for the system to be stable.

While many characteristics of interest can be estimated using the regenerative method, we shall restrict our attention to the waiting time of the $n^{\underline{th}}$ customer, W_n (time from arrival to commencement of service). For further discussion of the simulation of the GI/G/1 queue see Crane and Iglehart (1974b). To obtain a representation for the process $\{W_n : n \geq 0\}$ let $X_n = v_{n-1} - u_n$ and set $S_0 = 0$, $S_n = X_1 + \cdots + X_n$, $n \geq 1$. The following well-known recursive relationship exists for the W_n's:

$$W_0 = 0, \quad W_{n+1} = [W_n + X_{n+1}]^+, \quad n \geq 0 .$$

By induction, we also have

$$W_n = \max\{S_n - S_k : k = 0, 1, \ldots , n\}, \quad n \geq 0.$$

Using the strong Markov property of the process $\{S_n : n \geq 0\}$ it can be shown that there exists a sequence of integer-valued r.v.'s $\{T_k : k \geq 0\}$ such that $T_0 = 0$, $T_k < T_{k+1}$, and $W_{T_k} = 0$ with probability one. In other words, the customers numbered T_k

are those lucky fellows who arrive to find a free server and experience no waiting in the queue. The fact that there exists an infinite number of such customers in the GI/G/1 queue is a direct consequence of the assumption that $\rho < 1$ and the strong law of large numbers. Thus $\{W_n : n \geq 0\}$ is a regenerative process. If we let $\tau_k = T_k - T_{k-1}$, $k \geq 1$, then τ_k represents the number of customers served in the k^{th} busy period (b.p.) and they are numbered $\{T_{k-1}, T_{k-1} + 1, \ldots, T_k - 1\}$. Next define the sequence $\{Y_k : k \geq 1\}$ by

$$Y_k = \sum_{j=T_{k-1}}^{T_k - 1} W_j ,$$

the sum of the waiting times in the k^{th} b.p. Since the queue is stable for $\rho < 1$, we know that $W_n \Rightarrow W$ as $n \to \infty$ and wish to estimate $E\{W\}$.

We are interested in constructing controls, C_1, which are highly correlated with $Z_1 = Y_1 - r\tau_1$, but which are not so complex that we can not calculate their means. The controls we consider are of the form $C_1 = D_1 - \tau_1/\mu$, where the r.v. D_1 is an attempt to mimic Y_1. To compute the mean of C_1 we need, of course, to be able to compute $E\{\tau_1\}$. For M/G/1 queues we know that $E\{\tau_1\} = 1/(1-\rho)$. For GI/M/1 queues $E\{\tau_1\} = 1/(1-\delta)$, where δ is the root inside the unit circle of $z - E\{\exp[-\mu(1-z)u_1]\} = 0$, where u_1 is an interarrival time. If the queue in question is neither an M/G/1 or GI/M/1 queue, then the term τ_1/μ in C_1 will have to be approximated by another r.v. whose mean can

be computed. The factor $1/\mu$ multiplying τ_1 helps to make the variance reduction obtained independent of the scale parameter μ. Also it can be argued that the term τ_1/μ is then in the same units as D_1, namely, the unit of time. Several alternatives for D_1 are listed below, indexed by a superscript. They are

$$D_1^{(1)} = X_1^+ = \begin{cases} W_0 = 0 , & \tau_1 = 1 \\ W_0 + W_1 , & \tau_1 \geq 2; \end{cases}$$

$$D_1^{(2)} = \begin{cases} 0 & \tau_1 = 1 \\ X_1^+ + X_2^+ , & \tau_1 \geq 2; \end{cases}$$

$$= \begin{cases} W_0 = 0 , & \tau_1 = 1 \\ W_0 + W_1 , & \tau_1 = 2 \\ W_0 + W_1 + X_2^+ , & \tau_1 \geq 3; \end{cases}$$

$$D_1^{(3)} = (X_1^+ + X_2)^+ = W_2 ;$$

and

$$D_1^{(4)} = \begin{cases} 0 , & \tau_1 = 1 \\ X_1^+ + (X_1^+ + X_2)^+ , & \tau_2 \geq 2 \end{cases}$$

$$= \begin{cases} W_0 = 0 , & \tau_1 = 1 \\ W_0 + W_1 , & \tau_1 = 2 \\ W_0 + W_1 + W_2 , & \tau_1 \geq 3 \end{cases}$$

In general, it is more difficult to calculate $E\{D_1^{(i)}\}$ as i increases. On the other hand, as i increases $D_1^{(i)}$ comes closer to Y_1 and presumably results in more variance reduction. Let $C_1^{(i)} = D_1^{(i)} - \tau_1/\mu$, $i = 1,2,3$.

In our simulation runs for an M/M/1 queue the controls $C_1^{(2)}$, $C_1^{(3)}$, and $C_1^{(4)}$ were generally found to do much better than $C_1^{(1)}$. However, the more complicated controls, $C_1^{(3)}$ and $C_1^{(4)}$, gave little improvement over $C_1^{(2)}$. Thus with both variance reduction and ease of computation in mind, $C_1^{(2)}$ is our first choice. To compute $C_1^{(2)}$ for the M/M/1 queue first note that

$$P\{\tau_1 = 1\} = P\{v_0 < u_1\} = \int_0^\infty e^{-\lambda y} \mu e^{-\mu y}\, dy = (1 + \rho)^{-1} ,$$

$$E\{X_1 | X_1 > 0\} = \frac{\left(\frac{\rho}{1+\rho}\right) \int_0^\infty e^{-\mu y}\, dy}{P\{X_1 > 0\}} = \mu^{-1} ,$$

and

$$E\{X_1^+\} = E\{X_1 | X_1 > 0\} \cdot P\{X_1 > 0\} = \frac{\rho}{\mu(1+\rho)} .$$

Hence

$$E\{D_1^{(2)}\} = 0 + \frac{\rho}{1+\rho} E\{X_1^+ + X_2^+ | X_1^+ > 0\} = \frac{\rho}{1+\rho} \left[\frac{1}{\mu} + \frac{\rho}{\mu(1+\rho)}\right] ,$$

since X_1 and X_2 are independent. The expectation of $C_1^{(2)}$ is then given by

$$E\{C_1^{(2)}\} = \mu^{-1}\left[\frac{\rho}{1+\rho} + \left(\frac{\rho}{1+\rho}\right)^2 - \frac{1}{1-\rho}\right] .$$

Further discussion on the computation of the mean of controls such as $C_1^{(2)}$ for the GI/G/1 queue is contained in the Appendix. Computational results from our simulation runs are contained in Section 3.

2.3. Internal controls for Markov chains

The second example we consider is a discrete time Markov chain. Assume now that we are simulating an irreducible, aperiodic, positive recurrent Markov chain, $\underset{\sim}{X} = \{X_n : n \geq 0\}$, with the goal of estimating $r = E\{f(X)\}$. The controls we consider here are of the form

$$C_1 = \sum_{\ell=0}^{n_0 \wedge (\tau_1 - 1)} [f(X_\ell) - r_0] , \tag{2.6}$$

where r_0 is a guess of r, n_0 a fixed integer, and $n_0 \wedge (\tau_1 - 1)$ denotes the minimum of n_0 and $(\tau_1 - 1)$. Again our motivation for C_1 is to mimic

$$Z_1 = \sum_{\ell=0}^{\tau_1 - 1} [f(X_\ell) - r] = Y_1 - r\tau_1$$

Of course, we still must be able to compute $E\{C_1\}$ in order to implement the method of internal control variables. Let $\mathcal{E}$ be the state space of $\underset{\sim}{X}$, which we assume is finite, and P the

transition probability matrix. Furthermore, let g be the column vector with components $f(i) - r_0$ and ${}_jP$ be the matrix with elements

$$
{}_jp_{k\ell} = \begin{cases} p_{k\ell}, & \ell \neq j \\ 0, & \ell = j . \end{cases}
$$

Then if we let $X_0 = j$ and form regenerative cycles based on the times of return to state j, it can easily be shown, using the method developed in Hordijk, Iglehart, and Schassberger (1976), that

$$
E\{C_1\} = \left(\sum_{n=0}^{n_0} {}_jP^n g \right)_j , \tag{2.7}
$$

where j is the regenerative state and the subscript j means the jth component of the indicated vector. For two simple Markov chains the theoretical amount of variance reduction obtained using the control C_1 has been calculated for different regenerative states j and values n_0 and is contained in the next section. From (2.6) it is clear that the larger the integer n_0 selected the closer C_1 comes to duplicating Z_1. However, the larger n_0 the more difficult is the computation of $E\{C_1\}$ contained in (2.7).

3. Numerical Results

3.1. The M/M/1 queue

Even for the simple M/M/1 queue it is in general impossible to verify analytically what variance reduction will be obtained via the several internal controls listed in the previous section, or to get an idea of the magnitude of the effect. For something as simple as $C_1^{(1)}$ it is difficult to compute analytically the correlation between $C_1^{(1)}$ and Z_1 for the M/M/1 queue, and this is what is required in equation (2.5) to find the variance reduction.

Thus, we resorted to simulations to verify the amount of variance reduction obtained and the relative effectiveness of the various controls. In the final simulations all runs were performed on an IBM System 360/67 computer using the LLRANDOM package (Learmonth and Lewis (1973)) which generates random numbers according to the scheme given by Lewis, Goodman, and Miller (1969) and exponentially distributed random numbers using the Marsaglia "rectangle-wedge-tail" method. Tests of the random number generator are given in Learmonth and Lewis (1974).

Of the extensive simulation runs performed, we give here only a summary of the conclusions and two detailed tabulations in the case of the most suitable control.

(1) The controls $D_1^{(2)}$, $D_1^{(3)}$ and $D_1^{(4)}$ do much better generally than $D_1^{(1)}$, with little improvement over $D_1^{(2)}$ obtained by use of $D_1^{(3)}$ and $D_1^{(4)}$. We say generally because results vary unpredictably with λ and μ and their ratio ρ.

(2) Subtracting the number of customers served in a busy period generally improves the variance reduction. By making it dimensionally stable as in $C_1^{(2)}$ we obtain a "variance reduction" measured in terms of ratios of standard deviations, of approximately 70%, <u>uniformly</u> over λ and μ. This is roughly equivalent to halving the number of cycles (b.p.'s) that one must simulate; $(0.7)^2 \approx .5$. Much better reductions can be obtained for smaller ρ (i.e. $\rho = 0.25$) by specially designed controls; the point is that $C_1^{(2)}$ works even out at $\rho = 0.99$, where variance reduction is extremely important.

Table 1 shows results obtained by simulating an M/M/1 queue out to $n = 2000$ cycles with control $C_1^{(2)}$ and replicating the simulation either $m = 250$ or $m = 100$ times to estimate the variance of the estimators $\hat{r}(n)$, $\hat{r}_{CT}(n)$, and $\hat{r}_{CB}(n)$, where we drop the n for convenience. Here, we have specifically that

$$\hat{r}_{CB}(n) = \frac{\frac{1}{n}\sum_{k=1}^{n} Y_k}{\frac{1}{n}\sum_{k=1}^{n} [\tau_k + \beta(C_1^{(2)} - E\{C_1^{(2)}\})]} .$$

The estimated precision (standard deviations) of the estimates of $E(W)$ are given in brackets under the estimates.

The results in Table 1 are for $\rho = 0.5$ (and $m = 250$) and two values of μ and for $\rho = 0.99$ (and $m = 100$) for two values of μ commensurate with those given for

$\rho = 0.5$. The second, third and fourth columns in the Table give estimates of the correlations between the control and $Y_1 - \tau_1$ etc. from which the theoretical variance reduction can be computed. They are very close to the values given in the next to last column, the observed variance reduction, from which we deduce that estimating β_T and β_B affects the variance reduction only slightly. Overall there is negligible effect of different values of μ on the variance reduction obtained for the cases $\rho = 0.5$ and $\rho = 0.99$. For the results $\rho = 0.99$ given in Table 1, the variance reduction is 75%, which is about the same as for $\rho = 0.5$. For the case where the control is on the bottom, i.e. for $\bar{\tau}$, the variance reduction is not quite as good for control of $\bar{Y}$. Note too that the estimated values of $E\{W\}$ appear in some cases to be at least three or four standard deviations from the true mean. This is because the estimates $\hat{r}$, $\hat{r}_{CT}$ and $\hat{r}_{CB}$ can be seen from the 100 replications to be non-normal. In other words, for high $\rho(0.99)$, the simulation needs to be taken out further than 2000 cycles. In summary, the variance reduction obtained with the internal control is practically independent of the scale factor μ and the traffic intensity ρ.

3.2. Two simple Markov chains

We turn now to two simple Markov chains to give another illustration of the method of internal controls. The first chain is simply a recurrent random walk with state space $E = \{0,1,2,3,4\}$ and transition matrix

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 \\ 0 & 1/4 & 1/2 & 1/4 & 0 \\ 0 & 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix} .$$

Suppose we are simulating to estimate $r = E\{X\}$, which can be computed to have the value 2. Table 2 contains the theoretical values of $\sigma^2\{Z_1'\}/\sigma^2\{Z_1\}$ and $\rho(C_1, Z_1)$, where C_1 is given by (2.6) for several choices of n_0 and r_0. The entries in these tables were calculated using the methods in Hordijk, Iglehart, and Schassberger (1976). Note that the variance reduction is largest when $E\{\tau_1\}$ is smallest and n_0 is largest. This also yields the highest value of $\rho(C_1, Z_1)$ in keeping with our idea of having C_1 mimic Z_1. Not too much is lost in variance reduction as the guess, r_0, departs from the true value $r = 2$. A variance reduction of more than 50% is attainable, but depends on the state which is used as the regeneration point.

The second Markov chain represents an (s,S) inventory model; see Crane and Iglehart (1974) for further details. The state space $\mathcal{E} = \{6,7,\ldots,10\}$ and the P matrix is given by

$$P = \begin{bmatrix} 3/8 & 0 & 0 & 0 & 5/8 \\ 1/4 & 3/8 & 0 & 0 & 3/8 \\ 3/16 & 1/4 & 3/8 & 0 & 3/16 \\ 1/8 & 3/16 & 1/4 & 3/8 & 1/16 \\ 1/16 & 1/8 & 3/16 & 1/4 & 3/8 \end{bmatrix}$$

TABLE 1. Internally controlled ratio estimates of $E(W)$ in an M/M/1 queue. Change in correlations and variance reduction due to change of scale and change of traffic intensity ρ. Number of cycles, 2000; number of replications of the simulation, 250 for $\rho = 0.5$ and 100 for $\rho = 0.99$. The percentage column gives the percentage reduction in standard deviation of the estimate of $E(W)$ due to control, over the standard deviation of the uncontrolled estimated. Control C is $C_1^{(2)}$.

ρ	μ	$r_{C,Y_1-\tau_1}$	r_{Y_1,τ_1}	$r_{Y_1,C}$	Av $\hat{r}$ Av$(\hat{r}_{CT})$ Av$(\hat{r}_{CB})$	var$(\hat{r})$ var$(\hat{r}_{CT})$ var$(\hat{r}_{CB})$	% reduction	True E(W)
0.5	0.0500	0.73390	0.86665	-0.81991	10.00246 (0.05599)	0.78392	100%	10.000
					9.95958 (0.03932	0.36713	72%	10.000
					9.99513 (0.03876	0.37553	71%	10.000
0.5	5.000	0.73392	0.8666	-0.81993	0.10003 ((0.000560)	0.0000784	100%	0.0000
					0.09960 (0.000560)	0.0000369	72%	0.1000
					0.0995 (0.0003833)	0.0000377	71%	0.1000
0.99	0.10000	0.78524	0.95775	-0.97580	8.80116 (0.27489)	7.5566	100%	10.000
					9.20920 (0.20023)	4.0090	73%	10.000
					9.71925 (0.22302)	4.9739	81%	10.000
0.99	10.00000	0.78516	0.95773	-0.95779	880.255 (27.496)	75604.48	100%	1000.00
					921.032 (20.080)	40320.48	73%	1000.00
					971.474 (22.323)	49831.35	81%	1000.00

TABLE 2

Variance ratio $\sigma^2 Z_1' / \sigma^2\{Z_1\}$ (top row)

and Correlation $\rho(C_1, Z_1)$ (bottom row)

for the random walk model

		$r_0 = 1.5$		
Return State i	$E\{\tau_1\}$	$n_0=1$	$n_0=2$	$n_0=5$
0	8	0.97 -0.16	1.0 -0.01	0.91 0.30
1	4	0.82 0.43	0.73 0.52	0.57 0.65
2	4	0.62 0.62	0.51 0.70	0.40 0.78
3	4	0.94 0.25	0.91 0.31	0.84 0.40
4	8	0.97 -0.16	0.98 -0.16	0.99 -0.10
		$r_0 = 2$		
0	8	0.97 -0.16	0.98 -0.12	1.0 0.03
1	4	0.89 0.34	0.83 0.42	0.70 0.55
2	4	0.57 0.66	0.46 0.74	0.35 0.81
3	4	0.89 0.34	0.83 0.42	0.70 0.55
4	8	0.97 -0.16	0.98 -0.12	1.0 0.03

TABLE 3

Variance ratio $\sigma^2\{Z_1'\}/\sigma^2\{Z_1\}$ (top row)

and Correlation $\rho(C_1,Z_1)$ (bottom row)

for (s,S) Inventory Model

$r_0 = 8$				
Return State i	E_2	$n_0=1$	$n_0=2$	$n_0=4$
6	5.55	0.75 0.51	0.65 0.59	0.49 0.72
7	5.72	0.82 0.43	0.67 0.57	0.49 0.72
8	6.27	0.89 0.33	0.76 0.49	0.57 0.66
9	7.21	0.95 0.21	0.86 0.38	0.69 0.56
10	2.88	0.95 0.22	0.81 0.44	0.42 0.77
$r_0 = 9$				
6	5.55	0.75 0.51	0.75 0.50	0.78 0.47
7	5.72	0.88 0.34	0.80 0.45	0.73 0.52
8	6.27	0.93 0.27	0.83 0.41	0.71 0.54
9	7.21	0.95 0.23	0.86 0.37	0.72 0.53
10	2.88	0.79 0.46	0.57 0.66	0.24 0.87

Again we wish to estimate $E\{X\} = 8.297$. Table 3 contains the results analogous to those for the random walk (Table 2). The same general observations about $E\{\tau_1\}$ and n_0 made for Table 2 seem to hold here as well. Note however that the results are sensitive to the choice of r_0. Again a 50% reduction in variance can be attained.

4. Conclusions and Extensions

We have been able to obtain a 50% variance reduction using internal control variables, for the regenerative estimate of the limiting value of the mean waiting time in an M/M/1 queue. This reduction is obtained uniformly over all parameter values. It is fairly certain that the technique will work well with any GI/G/1 queue.

Internal control variables can be easily used with discrete time Markov chains. The examples used in this paper showed that a variance reduction of 50% is attainable. This figure is likely to vary widely with the particular Markov chain. Continuous time Markov chains and semi-Markov processes can be handled in the same way using the discrete time method of Hordijk, Iglehart and Schassberger (1976).

Another method of internal stratified sampling was also investigated. This method produced little overall variance reduction despite considerable effort.

Finally we note that it is possible to apply the idea of internal controls to the classical sample average estimate over a realization of fixed length m. Thus in estimating $E(W)$ in the GI/G/1 queue we have

$$\bar{W}(m) = \frac{1}{m} \sum_{j=0}^{m-1} W_j \, , \tag{4.1}$$

which may be written as

$$\bar{W}(m) = \frac{1}{m} \{ \sum_{k=1}^{N(m)} Y_k + Y'_{n(m)+1} \} , \tag{4.2}$$

where $N(m)$ is the number of completed busy periods in the queue in $[0,m]$, Y_k as before is the sum of the waiting times in the kth cycle, and $Y'_{N(m)+1}$ is the sum of the waiting times in the last, incomplete cycle.

A central limit theorem for $\bar{W}(m)$ holds for which the variance term is proportional to $E\{Z_k^2\}$, and which is now estimated from the random number $N(m)$ of cycles. Control is applied to the Y_k's in (4.2) just as it is applied in the ratio estimator $\hat{r}_{CT}(n)$. Call this estimator $\bar{W}_C(m)$. For the M/M/1 queue the variance reduction observed in the simulations was the same for all controls $C_1^{(i)}$ with the ratio estimator and with $\bar{W}_C(m)$.

The main reason for considering $\bar{W}_C(m)$ is that, while it loses the advantage of being an estimator using a fixed number of i.i.d. random variables, one can apply the classical <u>external</u> controls to $\bar{W}_C(m)$. Thus one could use the difference of the sum of the m arrival times and the m service times to control $\bar{W}_C(m)$. We have not yet tried this idea.

APPENDIX

To implement the internal control techniques for a GI/G/1 queue, certain theoretical parameters are required. In this appendix we shall indicate the values of these parameters in so far as they can be calculated. These values are either well-known or easily calculated. For a reference to the known formulas see Cohen (1969).

We begin with $E\{\tau_1\}$, the expected number of customers served in a busy period. For the general GI/G/1 queue recall that we let $X_n = v_{n-1} - u_n$ and $S_n = X_1 + \cdots + X_n$, for $n \geq 1$, with $S_0 = 0$. Then $\tau_1 = \inf\{n > 0 : S_n \leq 0\}$. The general expression for $E\{\tau_1\}$ is given by

$$E\{\tau_1\} = \exp\{\sum_{n=1}^{\infty} n^{-1} P[S_n > 0]\} ,$$

an impossible expression to evaluate in general. Another useful expression for $E\{\tau_1\}$ is

(A.1) $$E\{\tau_1\} = 1/P\{W = 0\} ,$$

where W is the stationary waiting time. In the special case of the M/G/1 queue, however, we have

$$E\{\tau_1\} = (1 - \rho)^{-1} .$$

Now for the queue GI/M/1 we can use (A.1) and the stationary distribution of the embedded Markov chain to conclude that

$$E\{\tau_1\} = (1 - \delta)^{-1} ,$$

where δ is the root inside the unit circle of

$$z - U\{\mu(1-z)\} = 0$$

with $U(s) = E\{e^{-su_1}\}$, Re $s \geq 0$, and where v_0 is an exponential(μ) r.v. It is easy to check that $\delta = \rho$ for M/M/1 queues. Daley (1975) has recently proposed the approximation to δ given by

$$\tilde{\delta} = a_1(1-\rho)^2 + 2(1-b^{-1})\rho + (2b^{-1}-1)\rho^2,$$

where $a_1 = P\{u_1=0\}$, $E\{u_1\} = 1$, and $b = E\{u_1^2\}$. This approximation gives good results in a number of examples calculated by Daley (1975) and may be useful for the purposes of this paper.

Next we turn to the computation of $P\{\tau_1=1\}$ and $P\{\tau_1=2\}$ For the GI/G/1 case we have

$$P\{\tau_1=1\} = P\{S_1 \leq 0\}$$

and

$$P\{\tau_1=2\} = P\{S_1 > 0, S_2 \leq 0\},$$

both of which can be worked out with a little effort. For the M/M/1 queue

$$P\{\tau_1=1\} = (1+\rho)^{-1}$$

and

$$P\{\tau_1=2\} = \rho(1+\rho)^{-3} .$$

For the M/G/1 queue

$$P\{\tau_1=1\} = V(\lambda) ,$$

where $V(\lambda) = E(e^{-\lambda v_0})$, while for the GI/M/I queue

$$P\{\tau_1=1\} = 1 = U(\mu) ,$$

where $U(s)$ is given above. For the $M/E_k/1$ and $E_k/M/1$ queues the value $P\{\tau_1=2\}$ can be calculated with some effort. As these expressions are cumbersome they shall be omitted.

Finally we give various partial expectations which are needed for computing the means of internal control $C_1^{(4)}$. Namely,

$$E\{S_1^+ + S_2^+, \tau_1 \geq 2\} = E\{S_1, S_1 > 0\} + E\{S_2^+, S_1 > 0\}$$

and

$$E\{\tau_1, \tau_1 \geq 2\} = E\{\tau_1\} - P\{\tau_1=1\}.$$

Here the symbol $E\{X,A\} = E\{X1_A\}$, where X is a r.v., A an event, and 1_A the indicator function of A. In the special case of the M/M/1 queue,

$$E\{S_1, S > 0\} = \rho/\{\mu(1+\rho)\} ,$$

$$E\{S_2^+, S_1 > 0\} = \left[2\left(\frac{\rho}{1+\rho}\right)^2 + \frac{\rho^2}{(1+\rho)^3}\right]\mu^{-1} ,$$

and

$$E\{\tau_1, \tau_1 \geq 2\} = 2\rho/\{(1-\rho)(1+\rho)\} .$$

REFERENCES

[1] COHEN, J. W. (1969). The Single Server Queue. North-Holland Publishing Co., Amsterdam.

[2] CRAMÉR, H. (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton.

[3] CRANE, M.A. and IGLEHART, D.L. (1974a). Simulating stable stochastic systems, I: General multi-server queues. J. Assoc. Comput. Mach. **21**, 103-113.

[4] CRANE, M.A. and IGLEHART, D.L. (1974b). Simulating stable stochastic systems,II: Markov chains. J. Assoc. Comput. Mach. **21**, 114-123.

[5] CRANE, M.A. and IGLEHART, D.L. (1975a). Simulating stable stochastic systems, III: Regenerative processes and discrete-event simulations. Operations Res. **23**, 33-45.

[5] CRANE, M.A. and IGLEHART, D.L. (1975b). Simulating stable stochastic systems, IV: Approximation techniques. Management Sci. **21**, 1215-1224.

[7] DALEY, D.J. (1975). A simple approximation for the root of an equation in GI/M/s queues. Technical Report, Statistics Department (IAS), The Australian National University, Canberra.

[8] HORDIJK, A., IGLEHART, D.L. and SCHASSBERGER (1976). Discrete time methods for simulating continuous time Markov chains. Adv. in Appl. Probability **8**, 772-788.

[9] IGLEHART, D.L. (1975). Simulating stable stochastic systems, V: Comparison of ratio estimators. Naval Res. Logist. Quart. **22**, 553-556.

[10] IGLEHART, D.L. (1976a). Simulating stable stochastic systems, VI: Quantile estimation. J. Assoc. Comput. Mach. **23**, 347-360.

[11] IGLEHART, D.L. (1976b). Simulating stable stochastic systems, VII: Selecting the best system. To appear in Management Sci.

[12] IGLEHART, D.L. (1977). The regenerative method for simulation analysis. To appear in Current Trends in Programming Methodology 3, Software Modelling and Its Impact on Performance. M. Chandy and R. Yeh, eds. Prentice-Hall Englewood Cliffs, New Jersey.

[13] LEARMONTH, G.P. and LEWIS, P.A.W. (1973). Naval Postgraduate School Random Number Generator Package LLRANDOM. Naval Postgraduate School Report NPS55Lw73061A.

[14] LEARMONTH, G.P. and LEWIS, P.A.W. (1974). Statistical tests of some widely used and recently proposed uniform number generators. Proc. Seventh Conference on Computer Science and Statistics, Western Periodicals Co., North Hollywood, Calif.

[15] LEWIS, P.A.W., GOODMAN, A.S. and MILLER, J.M. (1969). A pseudo-random number generator for the system/360. IBM Systems Journal 8, 136-146.